実世界情報システムプロジェクト
～ネオサイバネティックス研究グループ～
マイクロフォンアレイ計測

眞溪　歩

新領域創成科学研究科　複雑理工学専攻

概 要

マイクロフォンアレイ計測による音源分離の現実的な応用例として，特定の場所に存在する話者の音声抽出を検討する．問題設定として，2 音源 (内 1 音源は所定の位置にある) の 2 マイクロフォン計測による音源分離を取り挙げる．この条件設定では，最適化 (評価関数設定と収束演算) を伴わないバーチャルな ICA によって，音源分離が可能であることを示す.

1　はじめに

マイクロフォンアレイを用いた音源分離の研究には長い歴史がある．多くの問題設定は，音源に関する情報が互いに独立である以外は未知というブラインド設定が多い．この条件設定は Blind Source Separation(BSS) と呼ばれ，近年では独立成分分析 (ICA) が積極的に利用されている．しかし，工学的な応用を考えると，音源に対して何らかの事前情報を持つ場合も想定され，必ずしも BSS に当てはまらない場合も存在する．たとえば，特定位置に立っている話者の発話のみを抽出したい場合もあるだろう．その特定位置が複数あり，ハンズフリーで音声を抽出することが望まれる場合，固定のマイクロフォンアレイは天井などに設置して音源分離することが考えられる．そこで，ここでは，2 音源 (内 1 音源は所定の位置にある) の 2 マイクロフォン計測下での音源分離について検討する.

2　BSS による音源分離

本研究の問題設定とは異なるが，用いる要素技術が近しいため，BSS による音源分離について概説する.

2.1　混合モデル

ここでは，簡単のためと 3 節の問題設定と関係して，2 入力 2 出力システムとして混合モデルを説明する.

音源の信号を $s_p[n], (p = 1, 2)$，マイクロフォンで観測される信号を $x_m[n], (m = 1, 2)$，これらの組合せに対するインパルス応答を $h_{mp}[n]$ とする．インパルス応答は，この計測が行われる音源環境システムを表しているが，残響はいずれ無視される強度になるため FIR システム $h_{mp}[n], (n = 0, \cdots, N-1)$ であると考える．さて，これらの間には,

$$\begin{aligned} \begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} &= \begin{bmatrix} h_{11}[n] & h_{12}[n] \\ h_{21}[n] & h_{22}[n] \end{bmatrix} \otimes \begin{bmatrix} s_1[n] \\ s_2[n] \end{bmatrix} \\ \mathbf{x}[n] &= \mathbf{h}[n] \otimes \mathbf{s}[n] \end{aligned} \tag{1}$$

なる関係がある．ここで，$\otimes$ はたたみ込みを表す演算記号である．式 (1) を z 変換すると，たたみ込みは積に代わり，混合システム

$$\mathbf{X}(z) = \mathbf{H}(z)\,\mathbf{S}(z) \tag{2}$$

となる．

これに対し，瞬時混合は式 (3) によって与えられる．

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n], \quad (\mathbf{A} \in \mathbb{R}^{2\times 2}) \tag{3}$$

瞬時混合は，$h_{mp}[0]$ のみ非零の値を持つたたみ込み混合 (1) とも解釈される．このため，瞬時混合は，伝播遅延の存在するマイクロフォンアレイによる音声混合には当てはまらない．

2.2 FastICA

ICA では，音源 $s_1[n], s_2[n]$ が統計的に独立であることを仮定する．ここでは，ICA のアルゴリズムのひとつである FastICA を紹介する [1]. FastICA において，混合信号 $\mathbf{x}[n]$ はまずホワイトニング

$$\mathbf{x}'[n] = \mathbf{\Lambda}^{-1/2}\mathbf{V}^{\mathrm{T}}\mathbf{x}[n] \tag{4}$$

される．ここで，固有値分解

$$\left\langle \mathbf{x}[n]\,\mathbf{x}^{\mathrm{T}}[n] \right\rangle \;=\; \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$$

を利用している．ただし，$\langle\cdot\rangle$ は時間平均を表している．

ホワイトニングされた $\mathbf{x}'[n]$ は分散が規格化され，2次までの統計量が無個性となる．ここで，適当な非正規性を量る非線形関数 $g(\cdot)$ と射影ベクトル $\mathbf{u}$ を用意し，最適化

$$\langle g\left(\mathbf{u}^{\mathrm{T}}\mathbf{x}'[n]\right)\rangle \to \text{min. or max.} \tag{5}$$

を行う．求まった射影ベクトルを $\mathbf{u}_1$ とする．つぎに，独立であれば無相関なので，$\mathbf{u}_1$ の補空間で同様の最適化を行う．しかし，ここでは2次元空間なので，次の射影ベクトル $\mathbf{u}_2$ は $\mathbf{u}_1$ に直交する．これらの射影ベクトルを並べた行列 $[\mathbf{u}_1^{\mathrm{T}}\ \ \mathbf{u}_2^{\mathrm{T}}]$ にホワイトニング行列の逆行列 $\mathbf{V}\mathbf{\Lambda}^{1/2}$ を左から乗じれば，分離行列の転置行列 $\mathbf{B}^{\mathrm{T}}$ が求まる．この分離行列 $\mathbf{B}$ を用いて

$$\mathbf{y}[n] = \mathbf{B}\mathbf{x}[n]$$

とおくと，$y_q[n], (q = 1, 2)$ は互いに無相関で，かつそれぞれが非正規で個性的な確率密度関数を有する時系列となる．

2.3 たたみ込み混合の音源分離

ICA は，たたみ込み混合 (1) ではなく，瞬時混合 (3) の場合，非常に高性能な音源分離を実現する．瞬時混合に対する音源分離 $\mathbf{y}[n] = \mathbf{B}\mathbf{x}[n]$ では，分離行列 ($\mathbf{B} \in \mathbb{R}^{2\times 2}$) となり，4つのパラメータを定めればよい．一方，たたみ込み混合 (1) に対する音源分離 $\mathbf{y}[n] = \mathbf{w}[n] \otimes \mathbf{x}[n]$ では，分離行列 $\mathbf{w}[n] = [w_{qm}]$ は非常の多くのパラメータを有する 2×2 の FIR システムとなる．このため，ICA による音源分離は，瞬時混合に比べたたみ込み混合の方が著しく数理的に困難になる．この困難さは物理的には以下のように説明できる．音源とマイクロフォンとの距離，また音響環境での多重反射に応じて，音声はさまざまな遅れを持ってマイクロフォンに到達する．この遅れは各マイクロフォン間でさまざまな時間遅れとなる．音声信号は広帯域に分布するので，このさまざまな時間遅れは，周波数成分に応じてさらにさまざまな位相遅れを生む．いま，音源分離処理を離散・有限長で行うとすると，周波数も離散・有限長になる．このため，分離行列 $\mathbf{w}[n] \in \mathbb{R}^{2\times 2}$ としてしまうと，仮にある離散周波数に対しては音源分離できても，他の離散周波数ではできなくなる．瞬時混合では信号の振幅だけの情報で分離できるが，たたみ込み混合では振幅と位相の情報も必要になる．

このことから逆に，各周波数ごとの時系列に分離できれば，それぞれに瞬時混合の ICA を作用させればよいことが導かれる．

しかし，この戦略には以下に列挙する問題が存在する．第1に，時間周波数解析では，その不確定性から，時間分解能と周波数分解能は同時に高めることができない．ここで取り扱う問題では，位相差の観点から周波数分解能を下げることはできないので，時間分解能を下げることになる．たとえば，この問題の対策には，オーバーラップ時間の大きい短時間フーリエ変換が用いられている [2]．第2に，ICA の持つ分離信号の順序不定の問題がある．瞬時混合の混合行列と分離行列の積は $\mathbf{BA} = \mathbf{PD}$ となり，$\mathbf{P}, \mathbf{D}$ はそれぞれ順序行列，対角行列である．この $\mathbf{P}$ のために，ICA が音源を分離できたとしても源信号の順序まではわからな

いという問題である．このことは，各離散周波数ごとにICAを行う方式では深刻な問題を生じる．つまり，ある離散周波数の瞬時混合ICAの音源分離結果が，別の離散周波数の結果とどの組合せになるかが定まらない．たとえば，この問題の対策には，隣接する離散周波数間では時系列活動の相関が大きいなどの経験則が利用されている [2]．第3に，ICAの持つ分離信号のスケール不定の問題がある．これは上述の $\mathbf{D}$ のために，分離された信号の振幅が特定されないという問題である．たとえば，この問題の対策には，上述の分離された信号を分離行列の逆行列に作用させ模擬的な混合信号を作り，実測の混合信号と比較してスケールを定める方法 (Minimal Distortion Principle) が用いられている [2, 3]．

3 問題設定

式 (2) より，単純に $\mathbf{S}(z) = \mathbf{H}(z)^{-1}\mathbf{X}(z)$ が考えられる．ここで，

$$\mathbf{H}^{-1}(z) = \frac{\tilde{\mathbf{H}}(z)}{\det\{\mathbf{H}(z)\}} \quad (6)$$

$$\tilde{\mathbf{H}}(z) = \begin{bmatrix} H_{22}(z) & -H_{12}(z) \\ -H_{21}(z) & H_{11}(z) \end{bmatrix} \quad (7)$$

となる．さて，逆システム (6) が安定となるためには，$\det\{\mathbf{H}(z)\}$ の零点がすべて単位円内に存在しないといけない．しかし，この条件は非常に特殊な状況でしか成立せず，一般に逆システム (6) は不安定である．そこで，安定性に問題のないFIRシステムである余因子行列 (7) からなるシステムを利用し，

$$\mathbf{Y}(z) = \tilde{\mathbf{H}}(z)\mathbf{X}(z) \quad (8)$$

とおく．

図 1 に，混合システム (2) と余因子システム (8) の直接接続システムを示す．たとえば，$Y_2(z)$ に含まれる $S_1(z)$ 成分は，破線で示される2つの経路で伝達されるが，$-H_{11}(z)H_{21}(z) + H_{21}(z)H_{11}(z) = 0$ となり，相殺されてしまう．つまり，余因子システムによって音源分離は実現されている．ただし，$Y_2(z)$ に含まれる $S_2(z)$ 成分は，実経験した $H_{22}(z), H_{12}(z)$ に加えバーチャルに $H_{11}(z), -H_{21}(z)$ も経験するため，完全には復元されない．一般に $H_{mp}(z)$ はローパス特性を有するので，定性的に $y_2[n]$ は $s_2[n]$ に対し高音が抑制されて残響が長くなる．

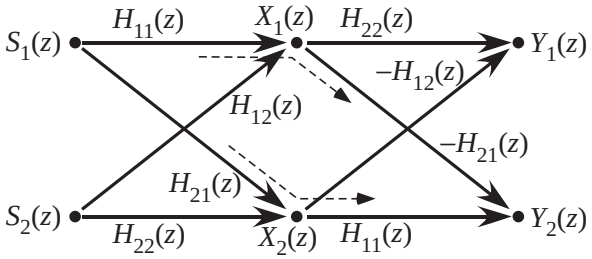

図 1: 混合システムと余因子システム

本研究の問題設定では，前述の2音源2マイクロフォンにおいて，音源1の位置情報は既知であるとする．具体的には，$h_{11}[n], h_{21}[n](H_{11}(z), H_{21}(z))$ は既知であるとする．また，音源 $s_1[n], s_2[n]$ は統計的に独立であるとする．物理環境に照らすと，片方の話者は所定の位置に立ち，他方はどこにいてもよく，これらの発話内容を分離することになる．式 (8) に照らすと，$H_{11}(z), H_{21}(z)$ の情報だけから $H_{21}(z), H_{22}(z)$ を推定することになる．

なお，周波数特性変化や残響の問題は，独立した問題としてここでは取り扱わない．これらの問題は，式 (6) における全極型システム $1/\det\{\mathrm{H}(z)\}$ を排除したことに起因している．この排除されたシステムは $Y_1(z), Y_2(z)$ に共通に作用している．このため，周波数特性変化や残響の問題への対策は，適当な復元フィルタを1つ設計すればよいこととなる．

4 提案手法

提案手法は，たたみ込み混合の分離法 [2] とFastICA[1] に基づいているが，3節の問題設定のためにICAはバーチャルにしか行わない．

まず，式 (1) における混合信号 $\mathbf{x}[n]$ を短時間

フーリエ変換する．このとき，後述するように切り出す幅はインパルス応答と同じ長さ N とし，適当なずらし幅でオーバーラップする区間が存在するように逐次切り出しを行う．切り出された順に番号 τ を与え，マイクロフォンごとに離散ベクトル時系列 $\mathbf{x}_m[n,\tau]$ を生成する．さらに，$\mathbf{x}_m[n,\tau]$ は，ハミング窓を乗じられた後，離散フーリエ変換され，複素離散時系列 $\mathbf{X}_m[k,\tau]$ となる．

この複素離散時系列 $\mathbf{X}_m[k,\tau]$ に対し，離散周波数 k を固定して変数 τ の時系列瞬時混合として FastICA を行うことを考える．まず，ホワイトニング (4) を行い，評価関数 (5) を最大または最小とする射影方向 $\mathbf{u}_1$ を求め，つぎに求まる射影方向 $\mathbf{u}_2$ を求める．これら 2 つの射影方向にホワイトニング行列の逆行列を乗じれば，そのどちらかが $[W_{21}[k] \;\; W_{22}[k]]^{\mathrm{T}} = [-H_{21}[k] \;\; H_{11}[k]]^{\mathrm{T}}$ になると仮定する．ここで，$H_{mp}[k] = \mathrm{DFT}\{h_{mp}[n]\}$(ただし，$\mathrm{DFT}\{\cdot\}$ は離散フーリエ変換) である．$[-H_{21}[k] \;\; H_{11}[k]]^{\mathrm{T}}$ は，3 節で問題設定したように，分離行列 (7) の既知であった第 2 行ある．つまり，この FastICA は，適切な評価関数 $g(\cdot)$ を有しており，分離行列 (7) を導いたと仮定するということである．この仮定の下では，自動的に $[W_{11}[k] \;\; W_{12}[k]]^{\mathrm{T}}$ も導かれる．図 2 にこの様子を示す．

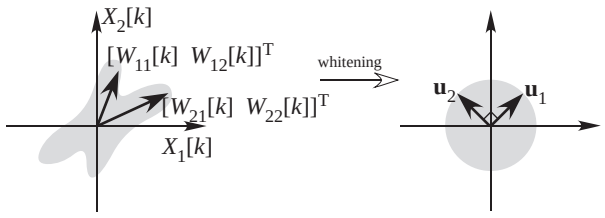

図 2: 離散周波数ごとの FastICA の様子

さて，逆に，既知である $[-H_{21}[k] \;\; H_{11}[k]]^{\mathrm{T}}$ にホワイトニング行列を乗じれば，ホワイトニングされた空間での射影方向 $\mathbf{u}_1$ が定まる．この射影方向 $\mathbf{u}_1$ に直交する方向 $\mathbf{u}_2$ は符合を除いて一意に決定される．最後に，$\mathbf{u}_2$ にホワイトニング行列の逆行列を乗じれば $[W_{11}[k] \;\; W_{12}[k]]^{\mathrm{T}}$，すなわち $[H_{22}[k] \;\; -H_{12}[k]]^{\mathrm{T}}$ の推定が求まる．

結局，複素時系列 $\mathbf{X}_m[k,\tau]$ に対する ICA が音源分離に効果的で，かつその結果求まる分離行列が第 2 行だけ既知の余因子行列 (7) であると仮定すれば，具体的な評価関数設定も収束演算も行うことなくバーチャルに ICA が施され，未知であった分離行列の第 1 行が定まることになる．また，このバーチャル ICA の利点には，順序不定性・スケール不定性の問題は起こらない，FastICA を複素時系列に拡張適用することを回避できることも挙げられる．

なお，このバーチャル ICA の一連の流れを無矛盾にするために，先の混合信号 $\mathbf{x}[n]$ の短時間フーリエ変換における切り出し幅を N とした．

5 まとめ

2 音源 (内 1 音源は所定の位置にある) の 2 マイクロフォン計測による音源分離に対し，提案手法であるバーチャル ICA が効果的であることを理論的に示した．現在，音源分離効果を実験的に確認中であるが，インパルス応答の正確な同定が重要である．この条件設定は，工学的応用としてもありえる範囲であり，本手法の利用価値も低くないと思われる．

参考文献

[1] A. Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis,” IEEE Transactions on Neural Networks, vol. 10, no. 3, pp. 626-634, 1999.

[2] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” IEEE Transactions on Speech and Audio Processing, vol. 12, no. 5, pp. 530-538, 2004.

[3] K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” Proc. ICA, pp. 722-727, 2001.